Diffusion Policies as an Expressive Policy Class for Offline Reinforcement Learning

Year: 2022 | Citations: 489 | Authors: Zhendong Wang, Jonathan J. Hunt, Mingyuan Zhou

Abstract

Offline reinforcement learning (RL), which aims to learn an optimal policy using a previously collected static dataset, is an important paradigm of RL. Standard RL methods often perform poorly in this regime due to the function approximation errors on out-of-distribution actions. While a variety of regularization methods have been proposed to mitigate this issue, they are often constrained by policy classes with limited expressiveness that can lead to highly suboptimal solutions. In this paper, we propose representing the policy as a diffusion model, a recent class of highly-expressive deep generative models. We introduce Diffusion Q-learning (Diffusion-QL) that utilizes a conditional diffusion model to represent the policy. In our approach, we learn an action-value function and we add a term maximizing action-values into the training loss of the conditional diffusion model, which results in a loss that seeks optimal actions that are near the behavior policy. We show the expressiveness of the diffusion model-based policy, and the coupling of the behavior cloning and policy improvement under the diffusion model both contribute to the outstanding performance of Diffusion-QL. We illustrate the superiority of our method compared to prior works in a simple 2D bandit example with a multimodal behavior policy. We then show that our method can achieve state-of-the-art performance on the majority of the D4RL benchmark tasks.